



Bayesian high-dimensional variable selection in non-linear mixed-effects models using the SAEM algorithm

Marion Naveau, Guillaume Kon Kam King, Laure Sansonnet, Maud Delattre

► To cite this version:

Marion Naveau, Guillaume Kon Kam King, Laure Sansonnet, Maud Delattre. Bayesian high-dimensional variable selection in non-linear mixed-effects models using the SAEM algorithm. Journée APPLIBUGS, Jun 2022, Lyon, France. hal-04247984

HAL Id: hal-04247984

<https://hal.inrae.fr/hal-04247984>

Submitted on 18 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Bayesian high-dimensional variable selection in non-linear mixed-effects models using the SAEM algorithm

Marion Naveau^{1,2}

M. Delattre², G. Kon Kam King², L. Sansonnet¹

¹Université Paris-Saclay, AgroParisTech, INRAE, UMR MIA Paris-Saclay

²Université Paris-Saclay, INRAE, MaIAGE

Journée AppliBUGS

10 Juin 2022

Table of contents

1. Introduction
2. Methodology
3. Simulation study
4. Application to a real dataset
5. Conclusion

1. Introduction

2. Methodology

- Prior specification
- Method
- Computation of the MAP

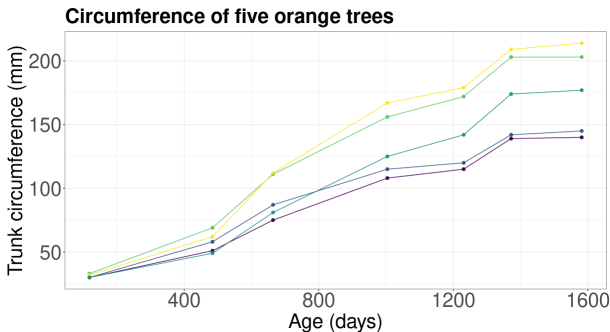
3. Simulation study

4. Application to a real dataset

5. Conclusion

Framework: repeated measurement data

- ❖ **Mixed-effects models:** analyse observations collected repeatedly on several individuals.



- ❖ Same overall behaviour but with individual variations.
- ❖ Non-linear growth.
- ❖ Are these variations due to known characteristics?
 - ▶ E.g.: growing conditions, genetic markers, ...

Non-linear mixed-effects model (NLMEM)

1) Description of intra-individual variability:

For all $i \in \{1, \dots, n\}$, $j \in \{1, \dots, J\}$,

$$y_{ij} = g(\varphi_i, \psi, t_{ij}) + \varepsilon_{ij}, \quad \varepsilon_{ij} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$$

- $y_{ij} \in \mathbb{R}$: response of individual i at time t_{ij} (**observation**).
- $\varphi_i \in \mathbb{R}$: individual parameter, **not observed**.
- $\psi \in \mathbb{R}^q$: fixed effects, **unknown**.
- g : **non-linear function** with respect to φ_i (**known**).

2) Description of inter-individual variability:

$$\varphi_i = \mu + {}^t\beta V_i + \xi_i, \quad \xi_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Gamma^2)$$

- $\mu \in \mathbb{R}$: intercept, **unknown**.
- $V_i \in \mathbb{R}^p$: covariates for individual i (**known**).
- $\beta = {}^t(\beta_1, \dots, \beta_p) \in \mathbb{R}^p$ covariate fixed effects vector, **unknown**.

Population parameters: $\theta = (\mu, \beta, \psi, \sigma^2, \Gamma^2)$

Variable selection

- ❖ **Aim**: identify the most relevant covariates to characterise inter-individual variability.
- ❖ **Active/Non-active covariates**: covariates that are actually influential/non-influential for the characteristic under consideration.
- ❖ Description of **inter-individual variability**:

$$\varphi_i = \mu + {}^t\beta V_i + \xi_i, \quad \xi_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Gamma^2)$$

- $\beta_\ell = 0 \iff$ covariate ℓ has **no effect** on parameter φ_i
- $\beta_\ell \neq 0 \iff$ covariate ℓ **gives some information** on parameter φ_i
- ❖ **Model selection**: variable selection $\iff$ model selection among all the possible supports of β :

$$S_\beta = \left\{ \ell \in \{1, \dots, p\} \mid \beta_\ell \neq 0 \right\}.$$

High-dimensional covariate selection in NLMEM

- ❖ **Goal:** identify the non-zero components of β .
- ❖ **Specificity of the problem:** $p \gg n$
- ❖ **Main difficulties:**
 - High-dimensional variable selection:
 - ▶ parsimonious estimation of β
 - Non-explicit likelihood
 - ▶ The φ_i 's are not observed (latent variables model)
 - ▶ g is non-linear

$$\begin{aligned} p(y; \theta) &= \int p(y|\varphi; \theta) p(\varphi; \theta) d\varphi = \prod_{i=1}^n \int p(y_i|\varphi_i; \theta) p(\varphi_i; \theta) d\varphi_i \\ &= C_{\sigma^2, \Gamma^2} \prod_{i=1}^n \int \exp \left(- \sum_{j=1}^J \frac{(y_{ij} - g(\varphi_i, \psi, t_{ij}))^2}{2\sigma^2} - \frac{(\varphi_i - \mu - {}^t\beta V_i)^2}{2\Gamma^2} \right) d\varphi_i \end{aligned}$$

State of the art for high-dimensional variable selection in mixed-effects models

❖ Frequentist framework:

- **LMEM**: both theoretical results and algorithmic developments for regularised methods (Schelldorfer et al., 2011; Fan and Li, 2012).
- **NLMEM**: algorithmic contribution (Ollier, 2021).

❖ Bayesian framework:

- **Linear regression (without random effects)**: $y_i = \alpha + \beta^t X_i + \epsilon_i$
theoretical and algorithmic developments using various **sparsity-inducing priors** (cf book Tadesse and Vannucci (2021)).
- **NLMEM**: (Lee, 2022) advocated the Bayesian approach for this model but this is only a review, without implementation, does not focus on the high-dimension.

Proposed approach

Association of a Bayesian *spike-and-slab* prior for variable selection with a stochastic version of the EM algorithm, called **MCMC-SAEM**, for inference.

1. Introduction

2. Methodology

- Prior specification
- Method
- Computation of the MAP

3. Simulation study

4. Application to a real dataset

5. Conclusion

Spike-and-slab prior for the coefficients of β

❖ Introduction of **latent variables** δ_ℓ , $1 \leq \ell \leq p$:

$$\delta_\ell = \begin{cases} 1 & \text{if covariate } \ell \text{ is to be included in the model,} \\ 0 & \text{otherwise.} \end{cases}$$

❖ **Spike-and-slab prior** on β (George and McCulloch, 1997):

$$\pi(\beta|\delta) = \mathcal{N}_p(0, \text{diag}((1 - \delta_\ell)\nu_0 + \delta_\ell\nu_1)), \quad 0 \leq \nu_0 < \nu_1 \text{ fixed,}$$

i.e. β_ℓ are independent and:

- $\beta_\ell | (\delta_\ell = 0) \sim \mathcal{N}(0, \nu_0)$: "spike" distribution, ν_0 small
- $\beta_\ell | (\delta_\ell = 1) \sim \mathcal{N}(0, \nu_1)$: "slab" distribution, ν_1 large

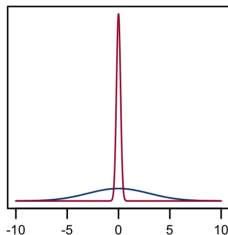
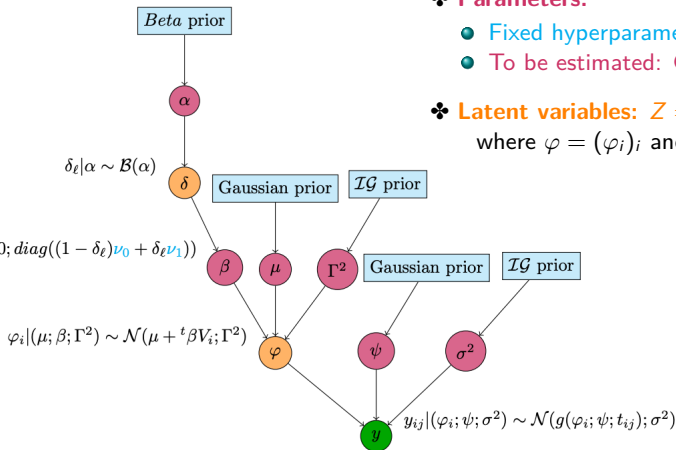


Figure: Spike-and-slab prior. Source: Deshpande et al. (2019)

Bayesian hierarchical model



❖ **Observations:** $y = (y_{ij})_{i,j}$

❖ **Parameters:**

● **Fixed hyperparameters:** $\nu_0, \nu_1, \dots$

● **To be estimated:** $\Theta = (\theta, \alpha)$

❖ **Latent variables:** $Z = (\varphi, \delta)$

where $\varphi = (\varphi_i)_i$ and $\delta = (\delta_\ell)_\ell$

Proposed method

Idea: explore different levels of sparsity in β by varying the value of ν_0 in a grid Δ .

1. **Creation of a model collection:** for each $\nu_0 \in \Delta$,

- Compute $\hat{\Theta}$ by a MCMC-SAEM algorithm (Kuhn and Lavielle, 2004):

$$\hat{\Theta}_{\nu_0}^{MAP} = \underset{\Theta \in \Lambda}{\operatorname{argmax}} \pi(\Theta|y)$$

- Estimate $\hat{\delta}$ (Ročková and George, 2014):

$$\hat{\delta} = \underset{\delta}{\operatorname{argmax}} P(\delta | \hat{\Theta}_{\nu_0}^{MAP}) \text{ such as } \hat{\delta}_\ell = 1 \iff \mathbb{P}(\delta_\ell = 1 | \hat{\Theta}_{\nu_0}^{MAP}) \geq 0.5$$

$$\iff \text{Define } \hat{S}_{\nu_0} = \left\{ \ell \in \{1, \dots, p\} \mid |(\hat{\beta}_{\nu_0}^{MAP})_\ell| \geq s_\beta(\nu_0, \nu_1, \hat{\alpha}_{\nu_0}^{MAP}) \right\}$$

2. **Select the "best" model** among $(\hat{S}_{\nu_0})_{\nu_0 \in \Delta}$ by a fast criterion, eBIC (Chen and Chen, 2008):

$$\hat{\nu}_0 = \underset{\nu_0 \in \Delta}{\operatorname{argmin}} \left\{ -2 \log(p(y; \hat{\theta}_{\nu_0}^{MLE})) + B_{\nu_0} \times \log(n) + 2 \log \left(\binom{p}{B_{\nu_0}} \right) \right\}$$

with B_{ν_0} : number of free parameters in the model $\hat{S}_{\nu_0}$.

3. **Return** $\hat{S}_{\hat{\nu}_0}$.

Spike-and-slab regularisation plot

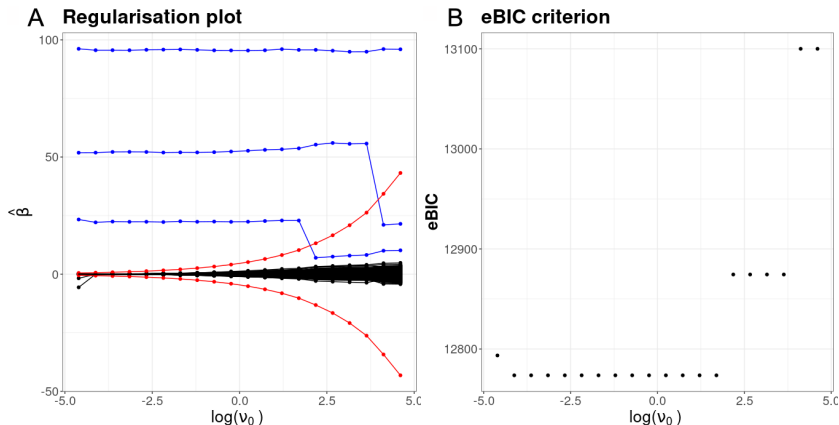


Figure: $n = 200$, $J = 10$, $p = 500$, $\Gamma^2 = 200$, $\sigma^2 = 30$, $\nu_1 = 12000$, $\mu = 1200$, $\beta = {}^t(100, 50, 20, 0, \dots, 0)$

Computing the MAP in a latent variables model

♣ Let's go back to the **first step** of the proposed method:

► Compute the MAP estimator of Θ

► **Goal:** maximise $\pi(\Theta|y) = \int_{\mathcal{Z}} \pi(\Theta, Z|y) dZ$ with

$$\pi(\Theta, Z|y) = \frac{p(y|\Theta, Z)p(\Theta, Z)}{\int_{\mathcal{Z}} \int_{\Lambda} p(y|\Theta, Z)p(\Theta, Z) d\Theta dZ}$$

► **Non-explicit integral**

EM algorithm

Reference: Dempster et al. (1977)

1. Initialisation: choose $\Theta^{(0)}$.
2. Iteration $k \geq 0$:
 - **E-step (Expectation):** compute

$$Q(\Theta|\Theta^{(k)}) = \mathbb{E}_{Z|(y, \Theta^{(k)})} \left[\log(\pi(\Theta, Z|y)) \middle| y, \Theta^{(k)} \right].$$

- **M-step (Maximisation):** compute

$$\Theta^{(k+1)} = \operatorname{argmax}_{\Theta \in \Lambda} Q(\Theta|\Theta^{(k)}).$$

3. $\hat{\Theta} = \Theta^{(K)}$, for K large enough.

SAEM and MCMC-SAEM algorithms

References: Delyon et al. (1999), Kuhn and Lavielle (2004)

1. Initialisation: choose $\Theta^{(0)}$ and $Q_0(\Theta) = 0$,
2. Iteration $k \geq 0$:
 - **S-step (Simulation):** simulate $Z^{(k)}$ using the result of one iteration of an MCMC procedure with $\pi(Z|y, \Theta^{(k)})$ for target distribution,
 - **SA-step (Stochastic Approximation):** compute an approximation of $Q(\Theta|\Theta^{(k)})$ according to:

$$Q_{k+1}(\Theta) = Q_k(\Theta) + \gamma_k(\log \pi(\Theta, Z^{(k)}|y) - Q_k(\Theta)),$$

- **M-step (Maximisation):** compute

$$\Theta^{(k+1)} = \operatorname{argmax}_{\Theta \in \Lambda} Q_{k+1}(\Theta),$$

3. $\hat{\Theta} = \Theta^{(K)}$, for K large enough,

where $(\gamma_k)_k$ a step sizes sequence decreasing towards 0 such that $\forall k$, $\gamma_k \in [0, 1]$, $\sum_k \gamma_k = \infty$ and $\sum_k \gamma_k^2 < \infty$.

Specifics in Spike-and-Slab-NLMEM

❖ Decomposition of Q :

$$\begin{aligned}
 Q(\Theta|\Theta^{(k)}) &= \mathbb{E}_{(\varphi, \delta)|(y, \Theta^{(k)})} [\log(\pi(\Theta, \varphi, \delta|y)) | y, \Theta^{(k)}] \\
 &= C + \underbrace{\mathbb{E}_{\varphi|y, \Theta^{(k)}} \left[\tilde{Q}_1(y, \varphi, \theta, \Theta^{(k)}) \middle| y, \Theta^{(k)} \right]}_{\text{non-explicit}} + \underbrace{\tilde{Q}_2(\alpha, \Theta^{(k)})}_{\text{explicit}}
 \end{aligned}$$

❖ M-step:

- ▶ θ and α estimated separately.
- ▶ $\hat{\alpha}$ updated as in an EM algorithm with $\tilde{Q}_2(\alpha, \Theta^{(k)})$.
- ▶ $\hat{\theta}$ updated via stochastic approximation of:

$$\mathbb{E}_{\varphi|y, \Theta^{(k)}} \left[\tilde{Q}_1(y, \varphi, \theta, \Theta^{(k)}) \middle| y, \Theta^{(k)} \right].$$

MCMC-SAEM algorithm in SSNLMEM

1. Initialisation: choose $\Theta^{(0)}$ and $Q_{1,0}(\theta) = 0$,
2. Iteration $k \geq 0$:
 - **S-step (Simulation)**: simulate $\varphi^{(k)}$ using the result of one iteration of an MCMC procedure with $\pi(\varphi|y, \Theta^{(k)})$ for target distribution,
 - **SA-step (Stochastic Approximation)**: compute

$$Q_{1,k+1}(\theta) = Q_{1,k}(\theta) + \gamma_k(\tilde{Q}_1(y, \varphi^{(k)}, \theta, \Theta^{(k)}) - Q_{1,k}(\theta)),$$

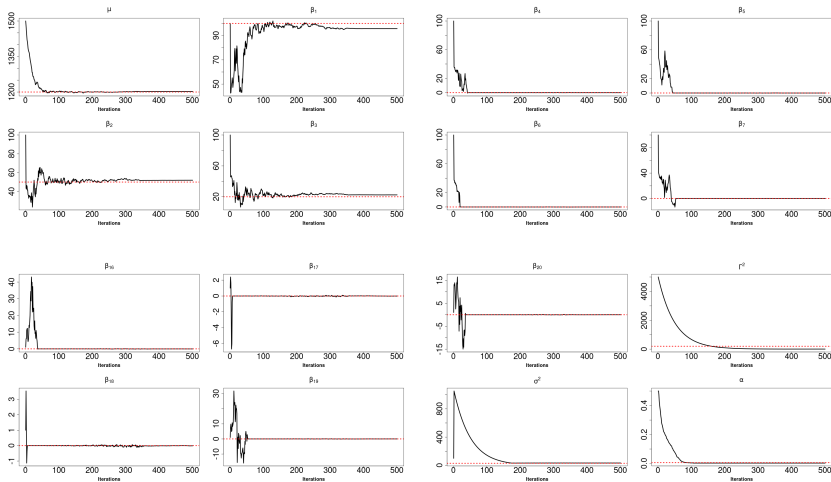
and $\tilde{Q}_2(\alpha, \Theta^{(k)})$,

- **M-step (Maximisation)**:

$$\theta^{(k+1)} = \operatorname{argmax}_{\theta \in \Lambda_\theta} Q_{1,k+1}(\theta) \text{ and } \alpha^{(k+1)} = \operatorname{argmax}_{\alpha \in [0,1]} \tilde{Q}_2(\alpha, \Theta^{(k)}),$$

3. $\hat{\Theta} = \Theta^{(K)}$, for K large enough,
 where $(\gamma_k)_k$ a step sizes sequence decreasing towards 0 such that $\forall k$, $\gamma_k \in [0, 1]$, $\sum_k \gamma_k = \infty$ and $\sum_k \gamma_k^2 < \infty$.

Convergence graphs



1. Introduction

2. Methodology

- Prior specification
- Method
- Computation of the MAP

3. Simulation study

4. Application to a real dataset

5. Conclusion

Logistic growth model

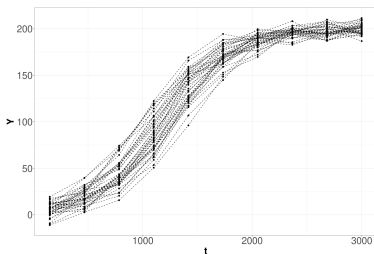


Figure: Simulated data

- Size of plant $i \in \{1, \dots, n\}$ at time t_{ij} , $j \in \{1, \dots, 10\}$:
 $y_{ij} = g(\varphi_i, \psi, t_{ij}) + \varepsilon_{ij}$, $\varepsilon_{ij} \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$ where:

$$g(\varphi_i, \psi, t_{ij}) = \frac{\psi_1}{1 + \exp\left(-\frac{t_{ij} - \varphi_i}{\psi_2}\right)}$$

$\psi = (\psi_1, \psi_2)$ fixed effects.

- φ_i : characteristic time
 $\varphi_i = \mu + {}^t\beta V_i + \xi_i$, $\xi_i \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \Gamma^2)$

$$\theta = (\mu, \beta, \psi, \sigma^2, \Gamma^2)$$

Simulation design

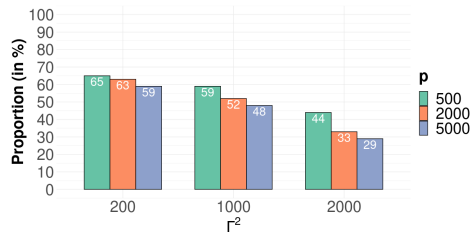
❖ Parameters:

- $n \in \{100, 200\}$ individuals,
- $p \in \{500, 2000, 5000\}$ simulated covariates according to $V_i \sim \mathcal{N}(0, \Sigma)$:
 - ▶ Scenario i.i.d.: $\Sigma = Id$ ▶ Correlated scenarios: $\Sigma \neq Id$
- $\beta = {}^t(100, 50, 20, 0, \dots, 0)$ covariate fixed effects vector,
- $\Gamma^2 \in \{200, 1000, 2000\}$ inter-individual variance,
- $\mu = 1200$, $\sigma^2 = 30$, $\psi = (\psi_1, \psi_2) = (200, 300)$.

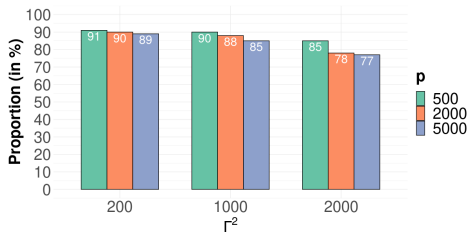
❖ Spike-and-slab hyperparameters:

- $\nu_1 = 12000$ slab variance,
 - $\log_{10}(\Delta) = \left\{ -2 + k \times \frac{4}{19}, k \in \{0, \dots, 19\} \right\}$ grid of ν_0 values.
- ▶ For each combination of (n, p, Γ^2) , the method is applied on 100 different simulated datasets.

Results for independent covariates



(a) $n = 100$



(b) $n = 200$

Figure: Empirical probability of correct model selection.

Summary of the results

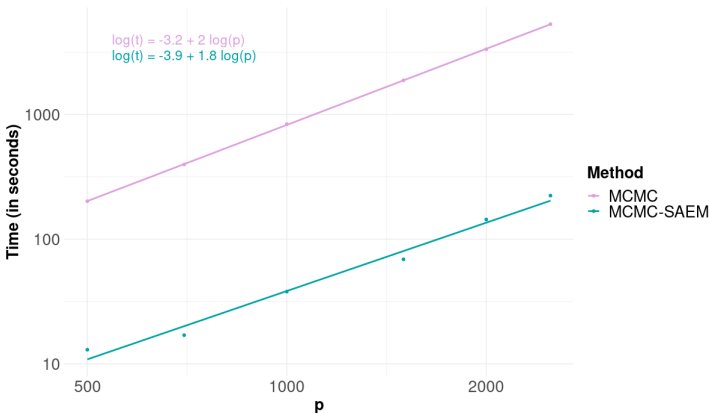
❖ Uncorrelated covariates $V_i \sim \mathcal{N}(0, I_p)$:

- Results improve as n increases.
- Degradation of results when p or Γ^2 increases.
- When the procedure fails, it is most often because it **under-selects**:
 - ▶ **"Cautious" approach**, few false positives!

❖ Correlated covariates $V_i \sim \mathcal{N}(0, \Sigma)$:

- **Fairly similar good performance.**
- **More false positives** and/or **false negatives** in some correlation scenarios:
 - ▶ + false positives: correlations between active and non-active covariates.
 - ▶ + false negatives: correlated active covariates.

Comparison with an MCMC implementation



NB: fast C++
adaptive MCMC
(Nimble) versus R
code

- Both methods have an execution time that grows **polynomially** with p .
- The proposed inference method can browse **grid of about 20 values** of ν_0 while adaptive MCMC explores a single value.

1. Introduction

2. Methodology

- Prior specification
- Method
- Computation of the MAP

3. Simulation study

4. Application to a real dataset

5. Conclusion

Presentation of the dataset

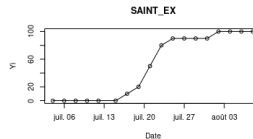
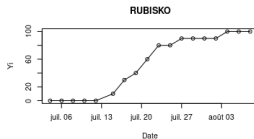
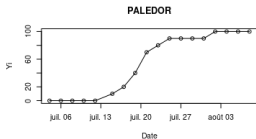
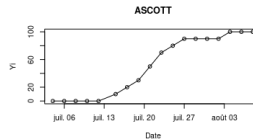
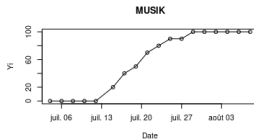
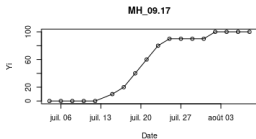
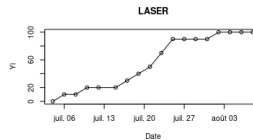
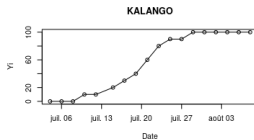
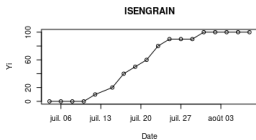
❖ Wheat leaf senescence data.

❖ **Panel:** $n = 216$ soft wheat **varieties** subjected to nitrogen stress, observed $J = 18$ times.

❖ Varieties **respond differently** to stress: for example, some of them tolerate stress better and senescence is delayed.

❖ **Aim:** select molecular markers, from among $p = 34838$ **markers**, which could be associated with this tolerance.

Data representation: percentage of desiccated leaves



⇒ Logistic growth

Modelling

$$\begin{cases} y_{ij} = g(\phi_i, t_{ij}) + \varepsilon_{ij} & , \varepsilon_{ij} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2), \text{ with } \phi_i = (\varphi_i, \psi_i) \in \mathbb{R}^2 \\ \varphi_i = \mu + {}^t\lambda v_i + {}^t\beta V_i + \xi_i & , \xi_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Gamma^2) \\ \psi_i = \eta + \omega_i & , \omega_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Omega^2) \end{cases}$$

where:

- $g(\phi_i, t_{ij}) = \frac{100}{1 + \exp\left(-\frac{t_{ij} - \varphi_i}{\psi_i}\right)}$,
- v_i : covariates not subject to selection, allows the inclusion of sub-populations in the model,
- V_i : molecular markers, subject to selection, which contains QTLs identified by biologists and markers associated with heading date which is highly correlated with φ_i .

$$\theta = (\mu, \lambda, \beta, \eta, \sigma^2, \Gamma^2, \Omega^2)$$

Data processing

- ❖ $p \gg n$: **ultra-high dimensional** problem.
- ❖ Molecular markers $\Rightarrow$ **strong correlations/collinearity** between covariates.

- ❖ Covariates have **few modalities**:

```
> table(nb_mod_cov)
```

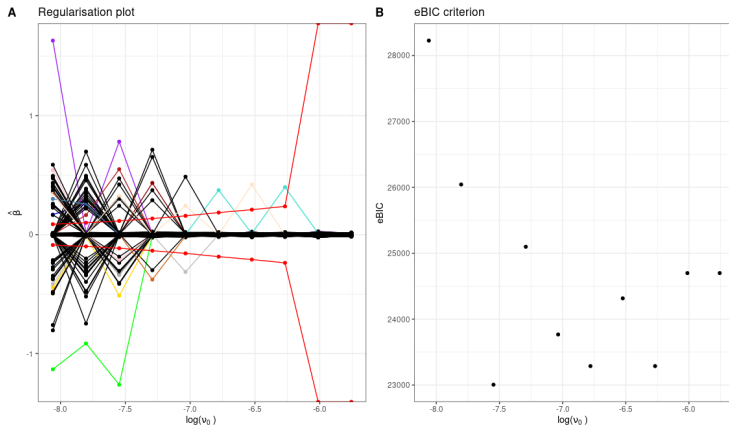
```
nb_mod_cov
```

	1	2	3	4
	45	9237	19712	5844

- ❖ With "too many" 0's or "too many" 1's for some covariates, we **remove**:
 - markers filled in the same way for all individuals,
 - markers entered as the exact opposite of another marker (marker1=1-marker2).
 - markers whose minimum and maximum modalities are not represented at least 10 times.
 - markers that have a correlation > 0.7 .

$$p = 6164$$

Results



- Selected support size: 20
- Number of covariates selected at least once along the grid: 90
- "Peak" structure could be explained by correlations between the covariates.

1. Introduction

2. Methodology

- Prior specification
- Method
- Computation of the MAP

3. Simulation study

4. Application to a real dataset

5. Conclusion

Conclusion and perspectives

❖ Summary:

- Development of an original method that combines SAEM and Bayesian variable selection.
- Very encouraging numerical results on simulated data.
- Faster method than a full MCMC implementation.

⇒ **Preprint:** Naveau and al. (2022). Bayesian high-dimensional covariate selection in non-linear mixed-effects models using the SAEM algorithm. [arXiv:2206.01012](https://arxiv.org/abs/2206.01012).

❖ Perspectives:

- Provide theoretical guarantees: selection consistency.
- Apply our method to a real dataset (in progress).
- Consider a multidimensional individual parameter.

Thanks for your attention!

References

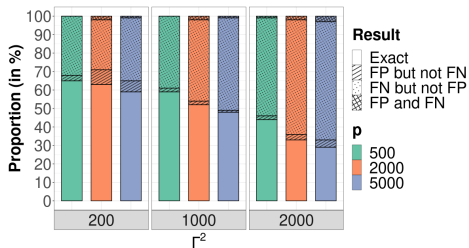
- Chen, J. and Chen, Z. (2008). Extended bayesian information criteria for model selection with large model spaces. *Biometrika*, 95(3):759–771.
- Delyon, B., Lavielle, M., and Moulines, E. (1999). Convergence of a stochastic approximation version of the em algorithm. *Annals of statistics*, pages 94–128.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.
- Deshpande, S. K., Ročková, V., and George, E. I. (2019). Simultaneous variable and covariance selection with the multivariate spike-and-slab lasso. *Journal of Computational and Graphical Statistics*, 28(4):921–931.
- Fan, Y. and Li, R. (2012). Variable selection in linear mixed effects models. *Annals of statistics*, 40(4):2043.
- George, E. I. and McCulloch, R. E. (1997). Approaches for bayesian variable selection. *Statistica sinica*, pages 339–373.
- Kuhn, E. and Lavielle, M. (2004). Coupling a stochastic approximation version of em with an mcmc procedure. *ESAIM: Probability and Statistics*, 8:115–131.
- Lee, S. Y. (2022). Bayesian Nonlinear Models for Repeated Measurement Data: An Overview, Implementation, and Applications. Publisher: Preprints.
- Ollier, E. (2021). Fast selection of nonlinear mixed effect models using penalized likelihood. *arXiv preprint arXiv:2103.01621*.
- Ročková, V. and George, E. I. (2014). Emvs: The em approach to bayesian variable selection. *Journal of the American Statistical Association*, 109(506):828–846.
- Schelldorfer, J., Bühlmann, P., and DE GEER, S. V. (2011). Estimation for high-dimensional linear mixed-effects models using 1-penalization. *Scandinavian Journal of Statistics*, 38(2):197–214.
- Tadesse, M. G. and Vannucci, M. (2021). Handbook of bayesian variable selection.

Specifics in Spike-and-Slab-NLMEM

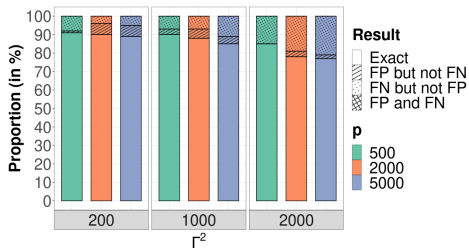
♣ Decomposition of Q :

$$\begin{aligned}
 Q(\Theta|\Theta^{(k)}) &= \mathbb{E}_{(\varphi, \delta)|(y, \Theta^{(k)})} [\log(\pi(\Theta, \varphi, \delta|y)) | y, \Theta^{(k)}] \\
 &= \mathbb{E}_{\varphi|(y, \Theta^{(k)})} \left[\mathbb{E}_{\delta|(\varphi, y, \Theta^{(k)})} \left[\log(\pi(\Theta, \varphi, \delta|y)) | \varphi, y, \Theta^{(k)} \right] \middle| y, \Theta^{(k)} \right] \\
 &= \mathbb{E}_{\varphi|(y, \Theta^{(k)})} \left[\tilde{Q}(y, \varphi, \Theta^{(k)}) \middle| y, \Theta^{(k)} \right] \\
 &= \underbrace{C + \mathbb{E}_{\varphi|y, \Theta^{(k)}} \left[\tilde{Q}_1(y, \varphi, \Theta^{(k)}) \middle| y, \Theta^{(k)} \right]}_{\text{non-explicit}} + \underbrace{\tilde{Q}_2(\alpha, \Theta^{(k)})}_{\text{explicit}}
 \end{aligned}$$

Results for uncorrelated covariates



(a) For $n = 100$

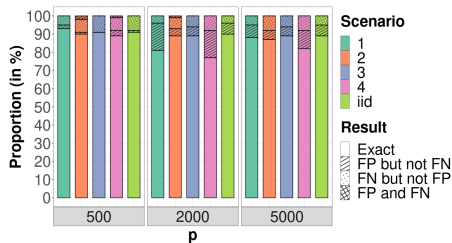


(b) For $n = 200$

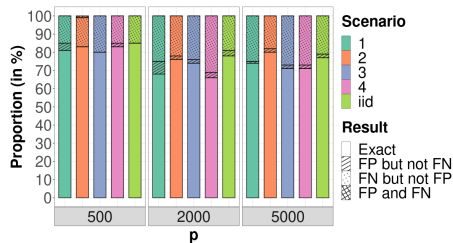
Correlated covariates $V_i \sim \mathcal{N}(0, \Sigma)$

Scenario	Σ
iid	I_p
1	$\left(\begin{array}{c c} I_3 & 0_{3,p-3} \\ \hline 0_{p-3,3} & (\rho_{\Sigma}^{ i-j })_{i,j \in \{4, \dots, p\}} \end{array} \right)$
2	$\left(\begin{array}{c c} I_3 & A \\ \hline {}^t A & I_{p-3} \end{array} \right), \text{ with } A = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ & & (\rho_{\Sigma}^{ 3-j })_{j \in \{4, \dots, p\}} \end{pmatrix}$
3	$\left(\begin{array}{c c} (\rho_{\Sigma}^{ i-j })_{i,j \in \{1, \dots, 3\}} & 0_{3,p-3} \\ \hline 0_{p-3,3} & I_{p-3} \end{array} \right)$
4	$(\rho_{\Sigma}^{ i-j })_{i,j \in \{1, \dots, p\}}$

Results for $\rho_{\Sigma} = 0.3$

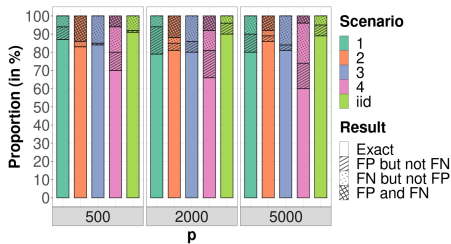


(c) For $\Gamma^2 = 200$

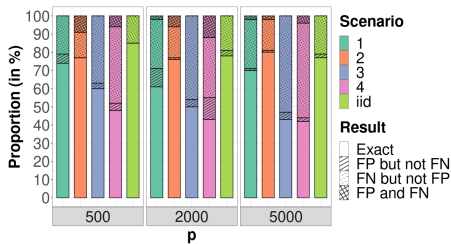


(d) For $\Gamma^2 = 2000$

Results for $\rho_{\Sigma} = 0.6$



(e) For $\Gamma^2 = 200$



(f) For $\Gamma^2 = 2000$